

Hierarchical clustering

- Hierarchical clustering is a widely used data analysis tool.
- The idea is to build a binary tree of the data that successively merges similar groups of points
- Visualizing this tree provides a useful summary of the data

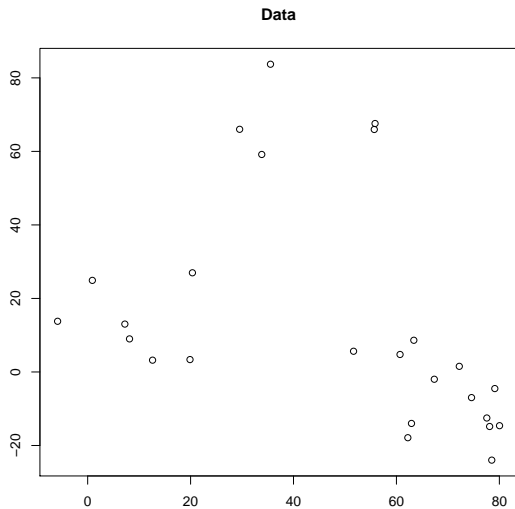
Hierarchical clustering vs. k -means

- Recall that k -means or k -medoids requires
 - A number of clusters k
 - An initial assignment of data to clusters
 - A distance measure between data $d(x_n, x_m)$
- Hierarchical clustering only requires a measure of similarity between *groups* of data points.

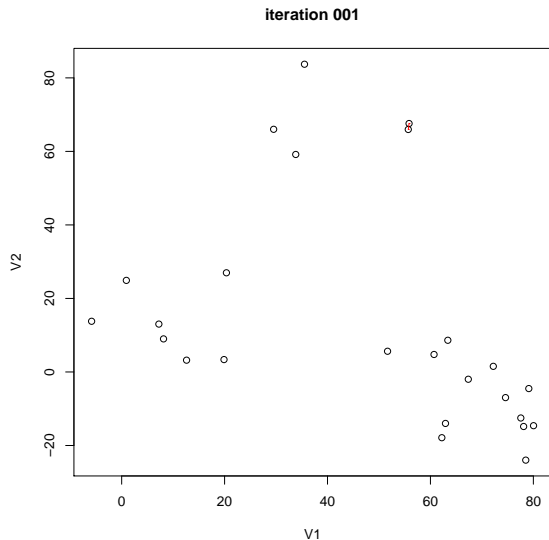
Agglomerative clustering

- We will talk about *agglomerative clustering*.
- Algorithm:
 - ① Place each data point into its own singleton group
 - ② Repeat: iteratively merge the two closest groups
 - ③ Until: all the data are merged into a single cluster

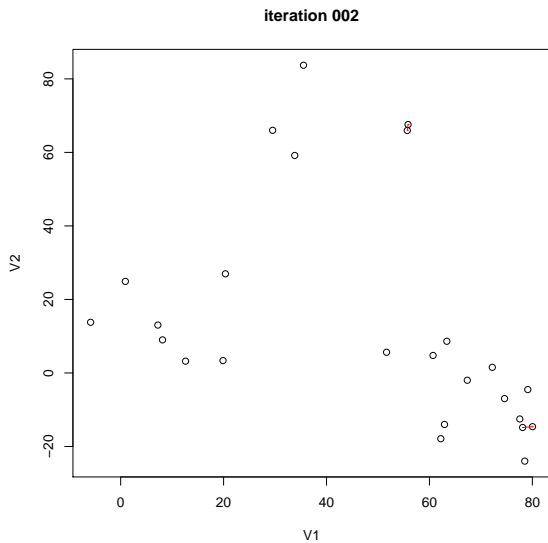
Example



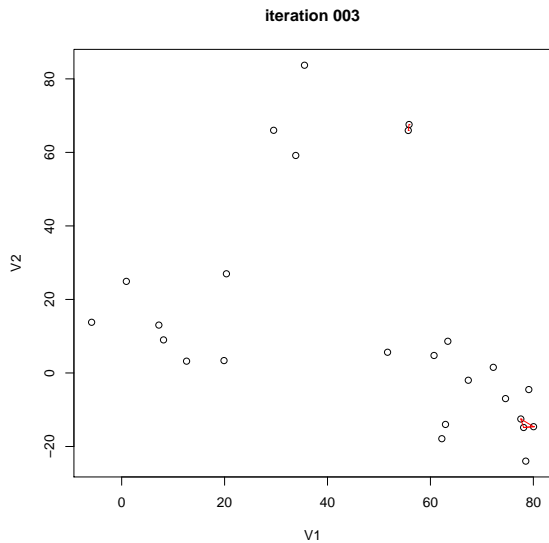
Example



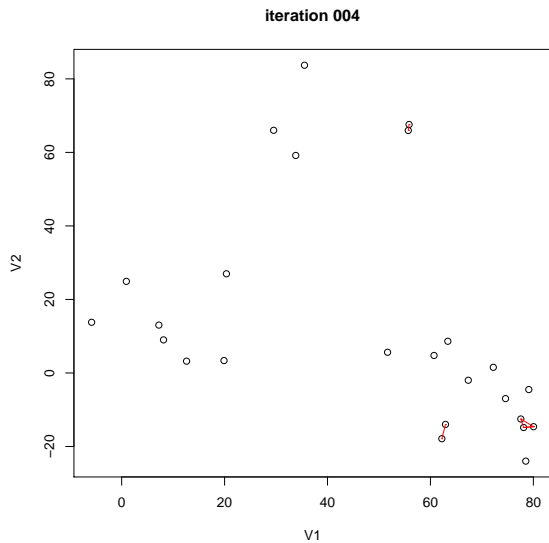
Example



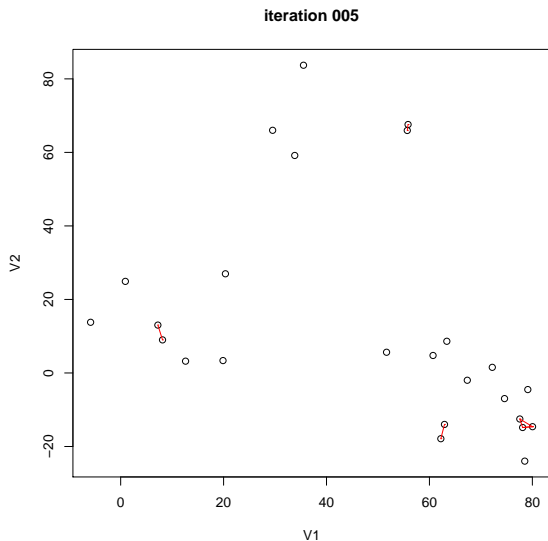
Example



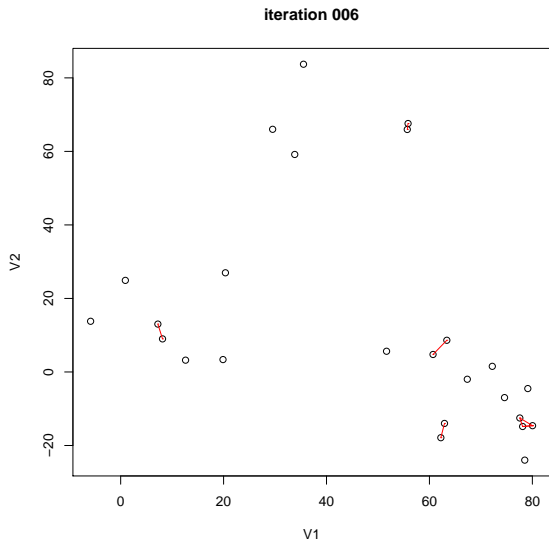
Example



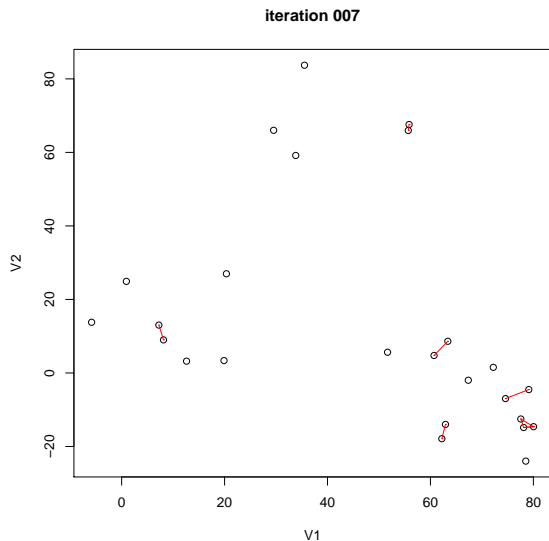
Example



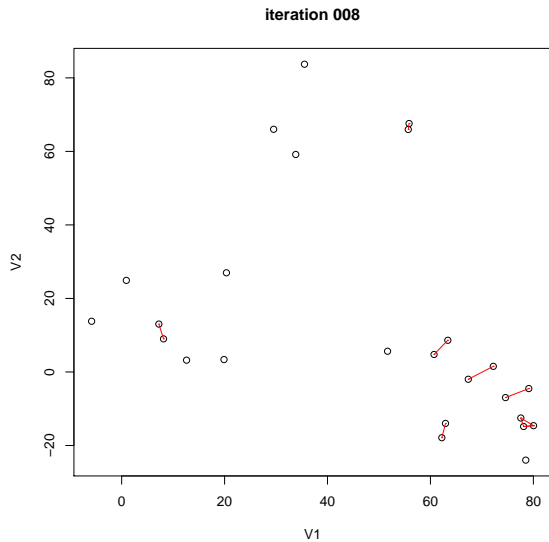
Example



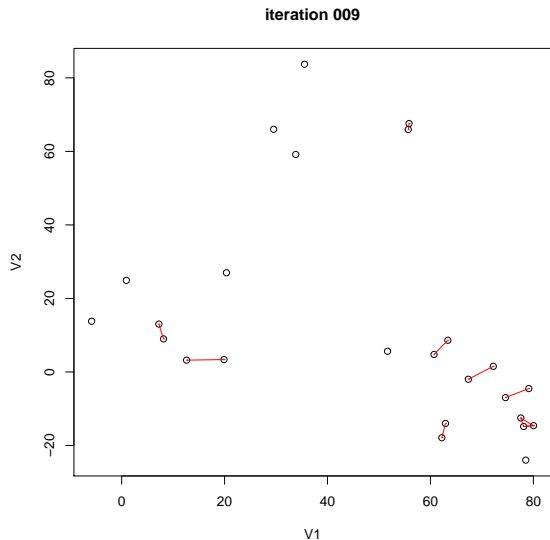
Example



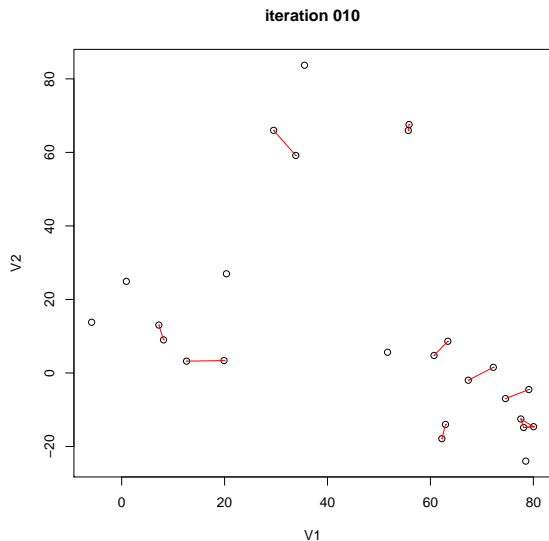
Example



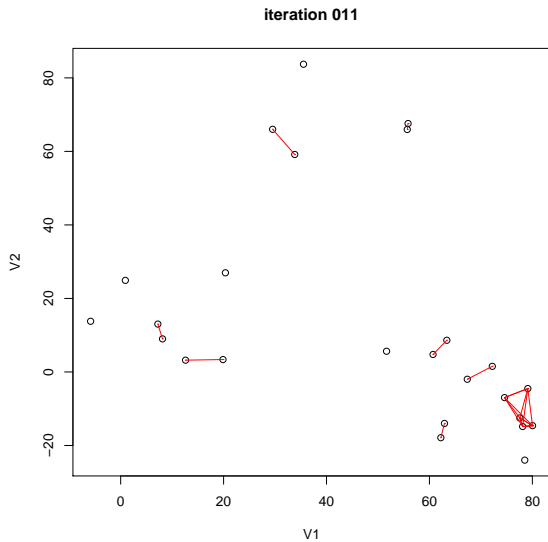
Example



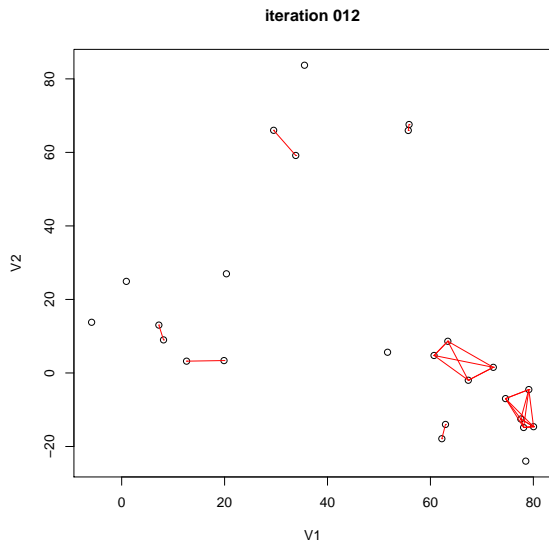
Example



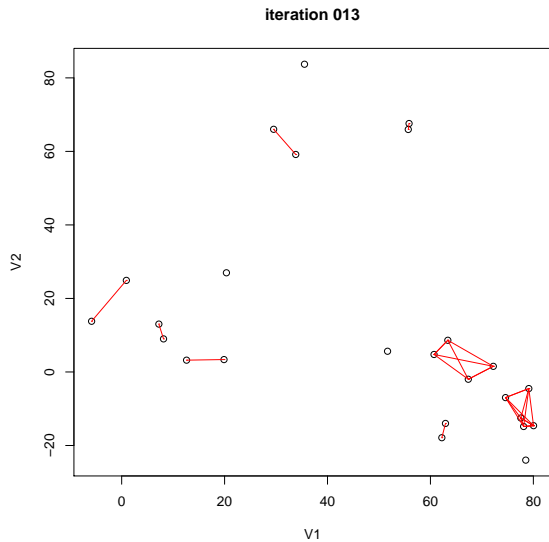
Example



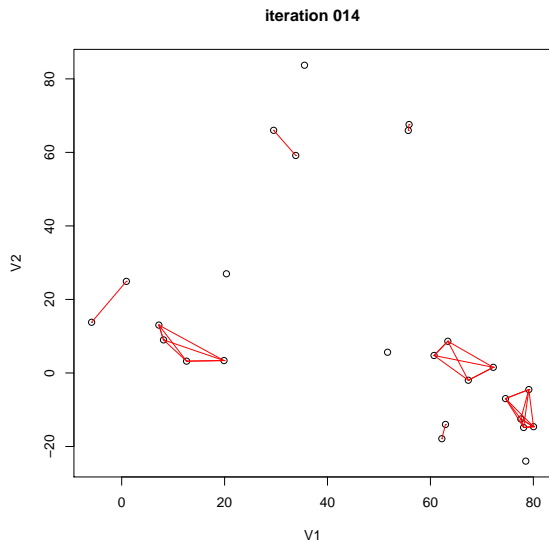
Example



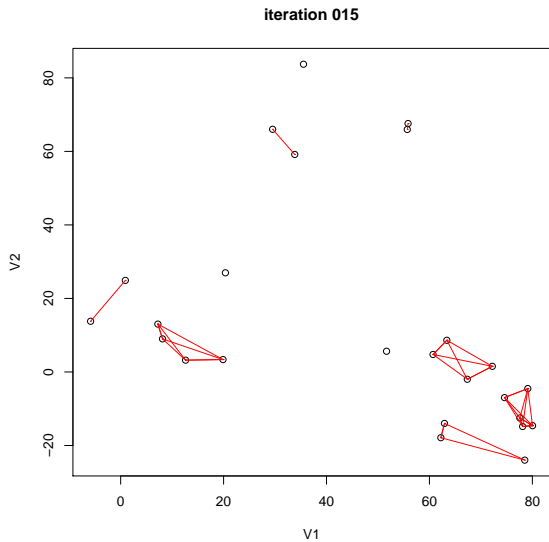
Example



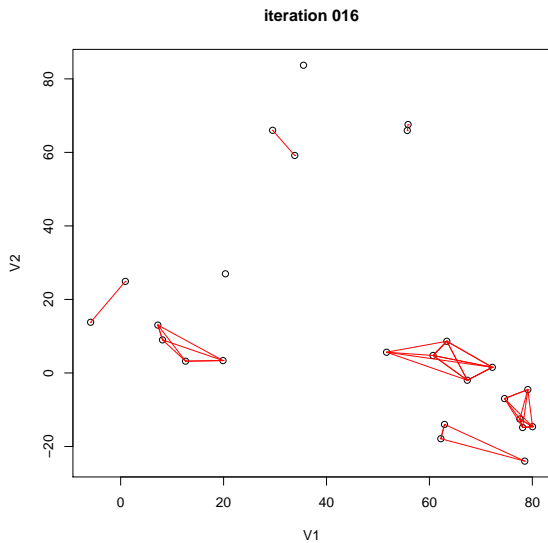
Example



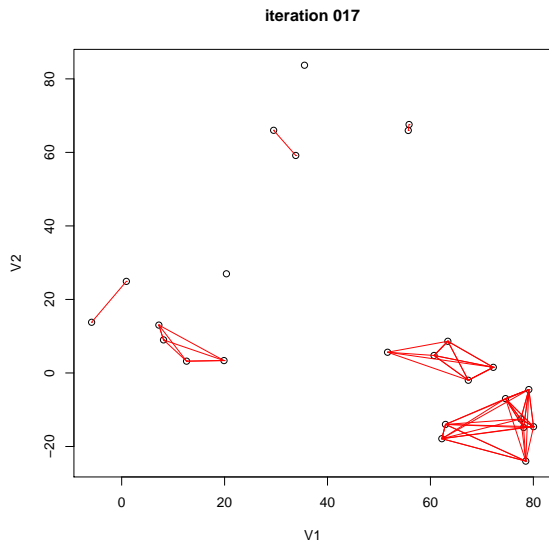
Example



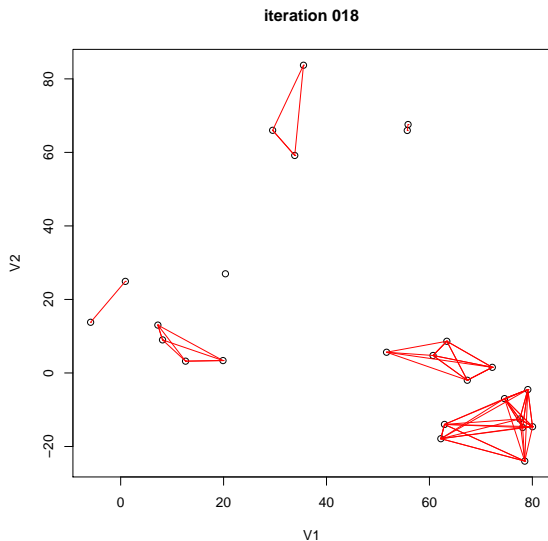
Example



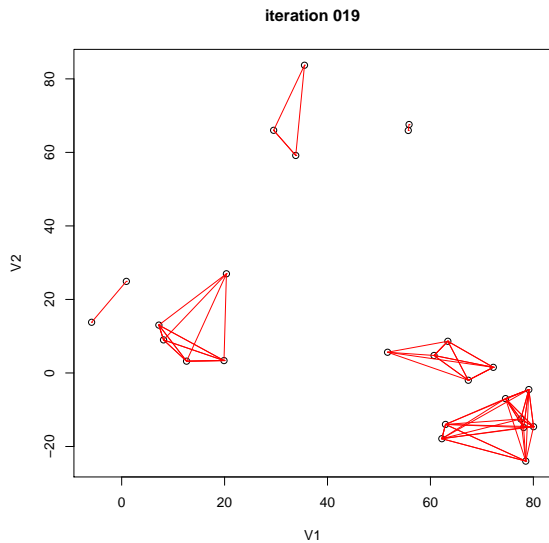
Example



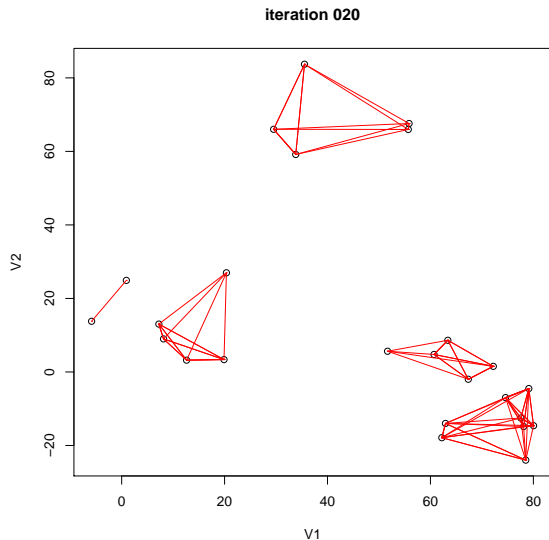
Example



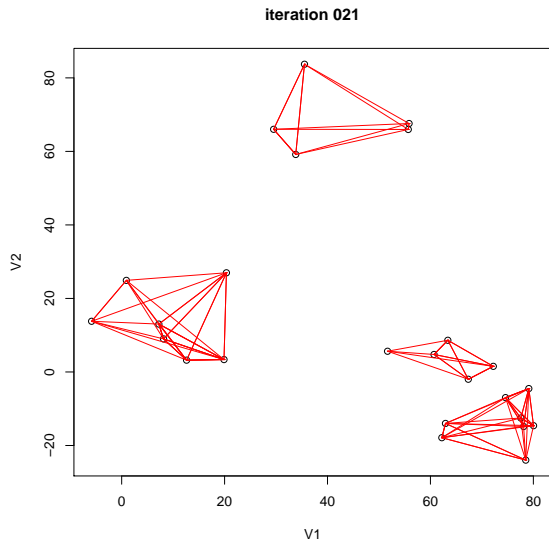
Example



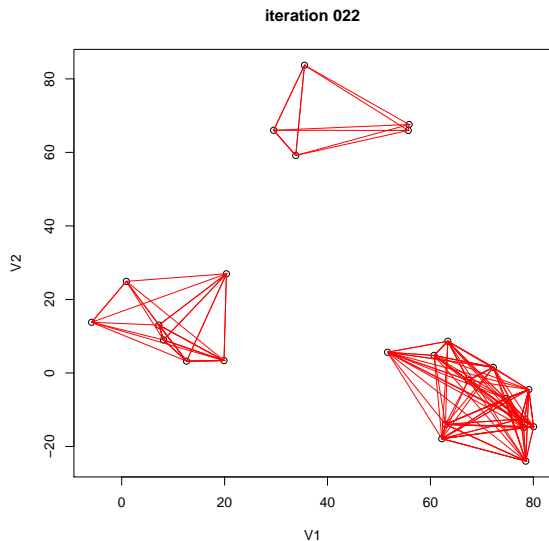
Example



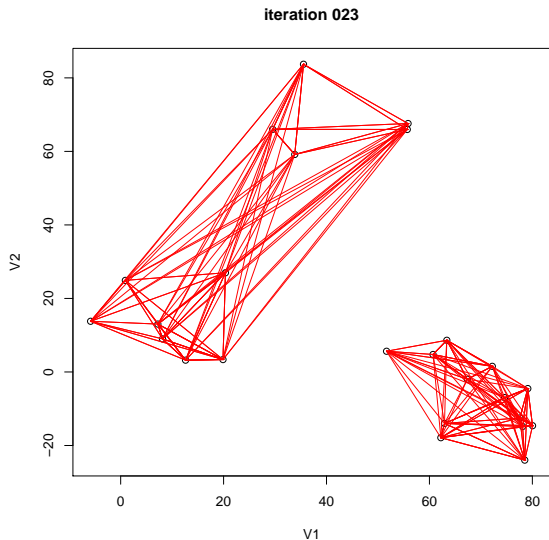
Example



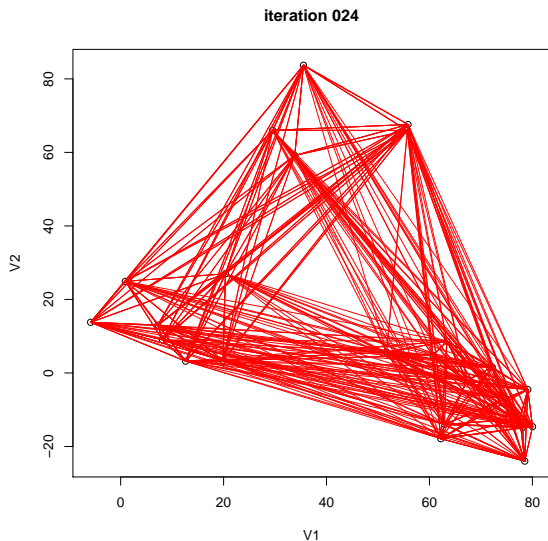
Example



Example



Example



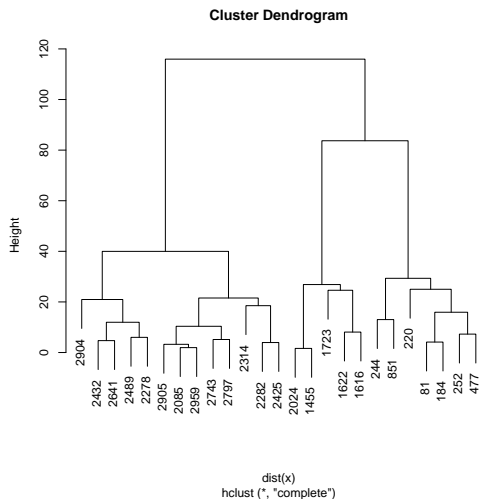
Agglomerative clustering

- Each level of the resulting tree is a segmentation of the data
- The algorithm results in a *sequence* of groupings
- It is up to the user to choose a "natural" clustering from this sequence

Dendrogram

- Agglomerative clustering is *monotonic*
 - The similarity between merged clusters is monotone decreasing with the level of the merge.
- *Dendrogram*: Plot each merge at the (negative) similarity between the two merged groups
- Provides an interpretable visualization of the algorithm and data
- Useful summarization tool, part of why hierarchical clustering is popular

Dendrogram of example data



Groups that merge at high values relative to the merger values of their subgroups are candidates for natural clusters. (Tibshirani et al., 2001)

Group similarity

- Given a distance measure between points, the user has many choices for how to define intergroup similarity.
- Three most popular choices
 - *Single-linkage*: the similarity of the closest pair

$$d_{SL}(G, H) = \min_{i \in G, j \in H} d_{i,j}$$

- *Complete linkage*: the similarity of the furthest pair

$$d_{CL}(G, H) = \max_{i \in G, j \in H} d_{i,j}$$

- *Group average*: the average similarity between groups

$$d_{GA} = \frac{1}{N_G N_H} \sum_{i \in G} \sum_{j \in H} d_{i,j}$$

Properties of intergroup similarity

- Single linkage can produce “chaining,” where a sequence of close observations in different groups cause early merges of those groups
- Complete linkage has the opposite problem. It might not merge close groups because of outlier members that are far apart.
- Group average represents a natural compromise, but depends on the scale of the similarities. Applying a monotone transformation to the similarities can change the results.

Caveats

- Hierarchical clustering should be treated with caution.
- Different decisions about group similarities can lead to vastly different dendrograms.
- The algorithm *imposes* a hierarchical structure on the data, even data for which such structure is not appropriate.